WHAT IS CLAIMED IS:

1. A method for automatically indexing and retrieving a multimedia event, comprising:

separating a multimedia data stream into audio, visual and text components;

segmenting the audio, visual and text components of the multimedia data stream based on semantic differences, wherein frame-level features are extracted from the segmented audio component are in a plurality of subbands;

identifying at least one target speaker using the audio and visual components;

identifying semantic boundaries of text for at least one of the identified target speakers to generate semantically coherent text blocks;

generating a summary of multimedia content based on the audio, visual and text components, the semantically coherent text blocks and the identified target speaker;

deriving a topic for each of the semantically coherent text blocks based on a set of topic category models; and

generating a multimedia description of the multimedia event based on the identified target speaker, the semantically coherent text blocks, the identified topic, and the generated summary.

2. The method of claim 1, further comprising:

automatically identifying a hierarchy of multimedia content types.

3. The method of claim 2, wherein the multimedia content types include at least one of speakers, anchors, interviews, correspondence reports, multimedia content segments, general news stories, topical news stories, news summaries, and commercials.

4. The method of claim 1, further comprising:

converting the multimedia data stream from an analog multimedia data stream to a digital multimedia data stream; and

compressing the digital multimedia data stream.

5. The method of claim 1, wherein the extracted audio features from the audio component further comprise clip level features.

6. The method of claim 1, wherein the multimedia event includes a news broadcast and the target speakers include news anchorpersons.

7. The method of claim 1, wherein the step of identifying at least one speaker includes the process of identifying using Gaussian Mixture Models.

8. The method of claim 1, wherein the generated multimedia description is represented by at least one of a text description, a video description and a story icon.

9. The method of claim 1, further comprising:

storing the generated multimedia descriptions in a database.

10. The method of claim 1, further comprising:

presenting the generated multimedia description to a user.

11. The method of claim 10, further comprising:

playing back the segment of the multimedia event corresponding to the generated multimedia description to the user.

12. The method of claim 1, wherein the plurality of subbands comprises three

subbands.

13. The method of claim 12, wherein the frame level features in the three subbands are at least one of volume, zero crossing rate, pitch period, frequency centroid, frequency bandwidth and energy ratios.

14. A system that automatically indexes and retrieves a multimedia event, comprising:

a multimedia data stream separation unit that separates a multimedia data stream into audio, visual and text components;

a data stream component segmentation unit that segments the audio, visual and text components of the multimedia data stream based on semantic differences;

a feature extraction unit that extracts audio features from the audio component and the audio features comprising a frame-level feature in a plurality of subbands;

a target speaker detection unit that identifies at least one target speaker using the audio and visual components;

a content segmentation unit that identifies semantic boundaries of text for at least one of the identified target speakers, to generate semantically coherent text blocks;

a summary generator that generates a summary of multimedia content based on the audio, visual and text components, the semantically coherent text blocks and the identified target speaker;

a topic categorization unit that derives a topic for each of the semantically coherent text blocks based on a set of topic category models; and

a multimedia description generator that generates a multimedia description of the multimedia event based on the identified target speaker, the semantically coherent text

blocks, the identified topic and the generated summary.

15. The system of claim 14, wherein the multimedia description generator automatically identifies a hierarchy of multimedia content types.

16. The system of claim 15, wherein the multimedia content types include at least one of speakers, anchors, interviews, correspondence reports, multimedia content segments, general news stories, topical news stories, news summaries, and commercials.

17. The system of claim 14, further comprising:

an analog-to-digital converter that converts the multimedia data stream from an analog multimedia data stream to a digital multimedia data stream; and

a compression unit that compresses the digital multimedia data stream.

18. The system of claim 14, wherein the multimedia event includes a news broadcast and the target speakers include news anchorpersons.

19. The system of claim 14, wherein the target speaker detection unit identifies at least one target speaker using Gaussian Mixture Models.

20. The system of claim 14, wherein the multimedia description generator generates one or more multimedia description that are represented by at least one of a text description, a video description and a story icon.

21. The system of claim 14, further comprising:

a database that stores the generated multimedia descriptions.

22. The system of claim 14, wherein the generated multimedia descriptions are retrieved from the database and presented to a user.

23. The system of claim 22, further comprising:

a playback device that plays back the segment of the multimedia event corresponding to the generated multimedia description to the user.

24. The system of claim 14, wherein the plurality of subbands comprises three subbands.

25. A terminal that displays the multimedia descriptions generated by the multimedia description generator of claim 1.